DOCUMENT RESUME

ED 093 866 SP 008 235

AUTHOR Scriven, Michael
TITLE Evaluating Program Effectiveness or, If the Program is Competency-Based, How Come the Evaluation Is Costing So Much?
PUB DATE Mar 74
NOTE 16p.

EDRS PRICE MF-$0.75 HC-$1.50 PLUS POSTAGE
DESCRIPTORS Cost Effectiveness; *Educational Programs; *Evaluation Criteria; *Evaluation Methods; Evaluation Techniques; *Performance Based Teacher Education; *Program Evaluation

ABSTRACT

The concern with competency-based education is not whether it is any different from the more traditional approaches but whether it is worth the considerable effort it involves. There are several aspects of a competency-based program to be considered in its evaluation. The first is whether or not there is a justifiable need for the specified competency. Secondly, the market, if any, must be identified. The testing of the program or specific competency must be considered to assure that the test population is comparable to the population for which the program is to be used. The program should be compared for effectiveness to the alternatives on the educational market. It must be tested over an extended period of time to determine its long-range impact. An evaluator of competency-based programs must also determine if there are any undesirable side effects of the program and whether the process itself has any damaging attributes. It is also necessary to determine whether the positive results of a field test are actually due to the program or to some outside factor. After evaluation of all the aspects of the program and its development, the program as a whole must be considered in light of its cost to see if it is a cost effective improvement for the schools. (HMD)

ED 093866

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

BEST COPY AVAILABLE

Scriven
3/74

EVALUATING PROGRAM EFFECTIVENESS
or, If the program is competency-based, how come the evaluation is costing so much?

1. Introduction

A CB program is presumably thoroughly developed by the time it is in full-scale operation, which presumably means that it gets its graduates up to design level of quality and quantity of achievements. (We pass lightly over the fact that it may do this by lowering the design level in the light of realistic field trials.) Here at last, one feels--or is it just that one hopes?--is a program that knows what it does and says what it knows. Here we have laid to rest the misleading advertising of inflated phrases and vague promises that pervade traditional education catalogues.

Apart from the moral and scientific charms of the CB approach long extolled in prose and poetry (somehow the poetry seems to creep back into the catalogs) there does seem to be at least one obvious pecuniary advantage about this innovation, namely the elimination of any need for fancy evaluation. CBE, the honest fellow, wears his--or is it her--credentials on its sleeve. They need no embellishment.

I wish I could add to your enjoyment of this evening's banquet by encouraging you to believe that the CB path is indeed rewarded in this way, but I propose instead to contribute to indigestion by stressing that the direct cost and quality of evaluation in CB programs that I know about is not even in the ballpark for adequacy. Not, I hasten to add, that this puts it in a poor

SP008 735

BEST COPY AVAILABLE

position compared to most educational programs. While it is true that the CB approach offers the possibility of a modest saving in one dimension of evaluation, this saving is usually not realized or not realized fully; and CB evaluation designs too often leave unanswered several fundamental questions about the merit of the program. My chief interest this evening is to outline the basic features of a satisfactory evaluation design.

To ensure a sweet aftertaste to this rather bitter post-coffee cordial, I shall add a word or two about the costs of evaluation , directed to the thesis that the median net cost, to the appropriate agency, of a thorough evaluation, is negative.

2The Main Problem

The main problem with the CB 'movement', as I see it today, is not--as the sideline cynics are always saying about innovations--that it does the same old things and just gives them a different name. There really is a different orientation and a different methodology as well as a different vocabulary. The point can be quite well-illustrated by reading the program for this meeting. It evidences no mere rechristening activity, but an effort-by many participants--to focus and explain their aims more clearly, an effort that definitely affects the content of at least some of the presentations.

The main problem, however, is still something pretty simple. It is the problem whether the total effect of the new orientation /and methodology is worth the very considerable effort it has involved, especially in light of the dubious justification of many of the competencies aimed at. The aim may well

be successful, but it's not at allclear it was sensible.

The analogy with the behavioural objectives and criterion-referenced testing movement naturally springs to mind, as I travel around the major metropolitan school districts talking to school principals and district or state staff, I believe I have noticed a steady deterioration in the interest in what I call the credentialling of objectives or criteria at the expense of a pro forma interest in BOs or criteria as such, or perhaps an interest in them for what they contribute to conceptualizing the educational activity, whether planning, performing, or reporting it. (Cf. also programming of texts)

Precisely the same problem-- as is no doubt obvious to most of you--faces the CB movement. It began with a no-nonsense orientation towards pragmatic teaching and evaluation, with emphasis on pay-off skills. But pretty soon we begin to find that the concept of a competency floats free of the original identification as an indubitable necessity for satisfactory job-performance, and is taken to include all sorts of juggling acts that someone or other has taken a fancy to. These are mere process competencies, and efforts to disseminate them in the absence of proof that they're important for pay-off performance represent exactly the kind of faddist, ill-grounded approach which the CB orientation was instigated to avoid. Only an uncompromising commitment to systematic and comprehensive evaluation of each program can avoid this kind of rot from undermining what could provide the foundations of a better approach to teacher education and to many other kinds of education. I'm delighted to see symptoms of concern with this point in some speaker summaries, e.g. N. L. Gage's reference to the need for first establishing desirability in teacher behaviour." It's

equally conspicuous by its absence in many other summaries. Perhaps that's just because they are summaries; we shall all find out soon.

3. The Basic Requirements for Good Evaluation

Let me try to provide, in checklist form, and as concisely as possible, an outline of the points that must be covered in the evaluation of a competency-based teacher education program. They derive from a general checklist, which was originally developed for the evaluation of educational products, but which has since been expanded to apply to programs and projects, (as well as to evaluators and proposals, etc.). This presentation will avoid the illustrative but slightly irresponsible anecdotes used in the Houston talk, where we were facing all the problems of after-dinner speeches, and it will omit the details of rating scales, etc., which are required in order to implement a checklist evaluation in practice. The intention is simply to play out the many dimensions which have to be explored in the course of evaluation, and which are not met by the mere waving of the magic wand on which are inscribed the words "competency-based".

3.1 Need

The first question, of course, is the one that I mentioned a little earlier--the question of the justification for this particular competency, set of competencies, programs committed to teaching us this set of competencies, etc. (For convenience, I'll focus on a component of a CBTE project, which is devoted to teaching a particular competency.) The problem of credentialling any successful training program is essentially the problem of hooking up its products to a needs assessment. Can we show that the need in question is a real need? Needs assessments are typically done in two completely spurious ways; either by using standardized tests and assuming that short falls on these tests represent need, or by doing a survey of the population that will be affected (or that is thought to be the population that will be affected) by the training program. The deficiencies with each should be

obvious, although both are provided as examples in current textbooks devoted in whole or in part to program evaluation. There is no need in the world to be performing at norm level on any standardized test. For that matter, there isn't any need to be performing at norm level on any criterion-referenced test, unless the criterion can be demonstrated to concern, e.g., survival or ability to cope with further experiences that will be leading one towards an adequate collection of survival skills, or etc. Given the differentials in rates and directions of maturation, it really wouldn't make much sense to say that somebody had a need to perform at norm level (mastery level) even on a criterion-referenced survival-connected curriculum except at the exit test.

The deficiencies in the usual survey approach are equally serious. People are very poor judges of what they really need (as opposed to what they want) where the issues they are passing judgment on are either technical or complex or emotional or novel. Since most of the areas where the surveys are taken concern issues that meet not just one but all of these conditions, they are staggeringly unreliable guides to need. I'm not suggesting for one moment that there is some omniscient entity around who can simply and reliably identify the "true needs" of populations of people. People may well be the best judge; but that doesn't mean that they're the best judge in the short run, when uninformed about the issue, and when torn by conflicting concerns over it. There are of course areas in teacher education where these problems are not serious; but if you start looking into the question of values clarification in social studies curricula, citizenship education, sex education, propaganda analysis, ecology, race relations, career choice, sex role, and a dozen other examples from the everyday activities of teachers in the school

classrooms of 1974, then you can readily see that the survey approach to needs assessment is likely to be unreliable. I'm not here to offer solutions--a solution to this problem requires some specification of the exact version of the problem that is facing one at the particular moment--but only to indicate the nature and dimensions of the problem, and I'll leave this point with the single thought that needs assessments of suitable degrees of reliability can indeed usually be done, but it takes a lot of work and thought to do it.

3.2 Market

What is needed is often unmarketable, and what is marketable is often not needed; once the distinction has been recognized, it's clear that we have to investigate market separately from need, or else we will be left with a training program for which the development money will have been wasted, since it cannot be disseminated to any significant number of the centers. NIE is in the process of rethinking its commitments in the direction of dissemination, and this is a most welcome move, since there has been a long tradition of assuming that dissemination will take care of itself, an assumption which shows an optimistic rather than a realistic nature.

3.3. Performance Data--True Field Trials?

Do we have a training program that really can deliver the promised competencies? Often we have one that has operated very successfully in an "intensive care" situation, guided by the inventor's hands and enthusiasm, but never tested out in the field, where it's got to sink or swim by itself. Where there have been field trials, they've often been semi-subsidized. with materials or assistance provided free of charge and that welcome little feeling of being part of the

in-group being supplied without any charge at all. As long as you can coast along on the Hawthorne effect, that will work splendidly; but it's a disservice to everybody to suppose that you can disregard the extent to which you owe your success to that handy little booster.

3.4 Performance Data--True Consumers?

Who have we looked at as the impacted population for our little educational package? Obviously, its primary impact (chronologically, at least) is on the teacher trainees who are going through the programs. A good evaluation doesn't rest on its laurels after looking at the first contact. After all, the justification of this little module is meant to lie, not in its beneficial effects on teachers, but in their beneficial effects on their students. So we normally would need to look into the question of whether we have data on the eventual student effects. The widening ripples do not cease at that point, however; do we have data on impact on future employers? What about impact on taxpayers? What about impact on other learning activities of the first level of students, i.e., the teacher trainees? What about special demands placed on the trainers themselves--have we investigated them as an impacted audience? The real consumers are a much less elite group than most evaluations suggest. And the news from the distant shores to which the effects finally spread is often not as good as the news from the points close to the point of first impact.

3.5 Performance Data--Critical Competitors

Useful evaluation is always--implicitly or explicitly--comparative. The real question for the person who is thinking about adopting or purchasing a competency-

based program is whether it is better than alternatives, not whether it is better than nothing at all. Hence, the evaluator sets up, or explores ERIC, for evidence about critical competitors. The choice is not between nothing at all, but between actual competing products or programs.

So, the choice is between a competency-based program and an existing program, or between the competency-based program and an alternative innovative program, e.g., apprenticeship systems with on-site workshops, etc.

Now, in the vanguard of new movements, there is always considerable resistance to having a new movement compared to the old one. The avant garde of the new curricula projects of the last decade were always talking about how inappropriate it was to compare them with the old ones. The legitimate part of this point was that it would have been unfair to compare them with the old ones, using only the old ones' criteria for the standards of comparison. However, it's not at all inappropriate to compare the new and the old on standards that embrace the best points of each. It is a matter of considerable significance to discover that Japanese children outperform their American age-peers, both on the Japanese curriculum tests and on the American curriculum tests. Whatever the explanation of this is, and it isn't necessarily a sign of the failure of the American school, the evaluator is making an extremely useful contribution when he or she identifies this difference. Now you can bet your life that competency-based approaches <u>always</u> have been compared to traditional approaches on at least one dimension--namely, cost. And the fact is that it's compared on a great many other dimensions too; the extent of disruption,

opposition by staff, etc., etc. It's a nice, comfortable little insulation move to suggest that the new is only to be evaluated on its own terms. But it's totally unrealistic, and it's illegitimate. It took many years for economists to discover that--to their embarrassment--a competitive market for refrigerators embraced not only refrigerators but dishwashers and electric stoves. That is, the rational consumer quite properly (and not irrationally) weighs the purchase of a new appliance against the purchase of other possible new appliances. Well, so it is with the competency-based approach. It has to earn its spurs in the free-for-all of the educational marketplace, not in some sequestered cloister.

3.6 Performance Data--Long-Term

We're very inclined to suppose that the termination of treatment is the appropriate time to measure the effects of treatment; and we thereby confuse logistical simplicity with methodological propriety. It's exceedingly inconvenient to have to segregate funds for use in long-term funds; but it's exceedingly unreliable to act as if somehow the results of such a follow-up will automatically be the same as the results of an evaluation at the termination of treatment.

3.7 Performance Data--Side Effects

Has there been a systematic, scientific, and sophisticated search for side effects? It has to be borne in mind that from the point of view of the evaluator, the goals of the program are really not a matter of great importance. They're certainly interesting for the archivist, for the developer, and for the funding agency. But what interests an evaluator is what the program actually did, judged by its congruence with the needs of the impacted population. It doesn't really matter much whether what it did was what it meant to do, or whether it failed to do what

it meant to do, but did something else instead, etc., etc. It's up to the evaluator to find out exactly what it did, and of course this means a great deal of fishing in murky waters where no guidance is provided by reference to goals statements. The search is not altogether without its own attractions; for the murky waters of goal statements, frequently inconsistent or incompatible with the goal statements of others associated with the project, or out of date, or based on factually false assumptions, etc., are by no means obviously more attractive than those wherein side effects lurk.

3.8 Performance Data--Process

It's quite wrong to use process as a reliable guide to outcome, but that doesn't mean that there aren't legitimate reasons for inspecting process in the course of evaluation. Four reasons are mentioned here:

a) the search for injustice. Whatever the outcomes, there are certain types of procedure that are inappropriate for moral reasons, and inspection of process must be made to see whether excess cruelty, inequitable methods of grading, etc., are important.

b) the search for joy. The second question is whether there is pleasure in the learning experience. There are other reasons for trying to make learning enjoyable, namely, that to do so is likely to increase the probability of learning taking place. But if all else fails, if learning gains are negible and if the cost is comparable, then we might as well take account of the question of whether the babies enjoy baby-sitting of type 1 more than they do type 2. I'd say that joy is a secondary end of education (because if it's joy you want, then for most people education has some overwhelmingly powerful critical competitors), but that doesn't mean it shouldn't be an aim.

c) characterization of the process. When the evaluation is done, somebody is going to have to label the package that was evaluated. Of course, it had a label

when it arrived at the evaluator's doorstep, but that label is as often as not extremely misleading; and one of the tasks of the evaluator is to decide whether this indeed deserves the title of "competency-based" training or not.

(d) causation. Inspection of process often provides valuable clues to determine the responsible agent for any changes that are observed. In particular, it may be necessary for you to inspect process in order to settle the question of whether the alleged cause was the real cause of the changes, in situations where control groups and baseline data are impossible.

3.9 Performance Data--Causation

Expanding on the last thought of the previous subsection, we have to face the fact that almost every evaluation constitutes an investigation of a causal claim, i.e., the claim that such-and-such a procedure had such-and-such (allegedly beneficial) effects. /This/ is one of the reasons why evaluation research is by no means an easy type of research. Quite often in teacher education, the amount of learning that takes place from peers is distinctly comparable to that which occurs from the training process itself; and the same applies to "natural" learning from field experiences as opposed to the contribution of the supervisor.

3.10 Performance Data--Statistical Significance

It is scarcely necessary to mention that evaluation frequently requires some rather elaborate statistical investigations, in order to determine the true significance of an apparent effect; or indeed in order to identify any effect whatsoever.

3.11 Educational Significance

When we are in possession of reports on the first ten checkpoints, we are in a position to tie them together and come up with a conclusion about educational significance. All too often, educational significance is identified with statistical significance of the difference between the treatment group and the control group. But of course that difference may be due to the use of a very large number of subjects (which magnifies trivial effects), to invalid tests, or to a dozen other causes, none of which represent any contribution to education. Checkpoint 11 is thus a summing-up checkpoint, where the results of detailed item analysis may come in, where investigations of instrument reliability may come in, and where it is frequently necessary to refer back to one's needs assessment data, in the light of one's refined characterization of the product, to see whether in fact the product as it turns out to be really meets the need that it was supposed to meet. Where the product--as in our case--is a competency-based training module, this kind of worry translates itself into worries about whether one can view the little skill that was acquired as generalizable to the many other similar skills that will be needed in the classroom.

3.12 Cost Effectiveness

We now have to look into questions of cost, and a very complicated business that is indeed. The use of a C.P.A. is not enough; the use of an economist is not enough; we usually have to use an economist with special expertise in the field of school economics; and a C.P.A. with special expertise in the various types of bookkeeping used by schools at primary, secondary, and tertiary levels.

When we have adequate data on the real, the opportunity, the direct and the indirect, the discounted cash flow, and other costs and analyses, then we can combine them

with the results of checkpoint 11 to yield either a cost-benefit analysis (at least) or a cost-effectiveness analysis (in most cases where we find reasonably simple dimensioning) is possible.

9.13 Continuing Support

This is a funny little item, as befits its number and its location on the checklist. But it's something of the sting in the tail of the checklist, too. It concerns the question of avoiding orphans, and when we start looking at orphans in the educational marketplace, we are much struck by the fact that the little entities that everybody wanted to be heir apparents to on one day are often entirely orphaned the next. Faddism is the great enemy of educational progress, and in order to avoid faddism, we need, in the course of a serious evaluation, to look at the question of sustained support and improvement in a given educational program, project, or product. Will Sony stop manufacturing video cassette recorders in the near future, just after we've purchased a set of them for the school? Will the behavioral objectives movement indeed be steadily amplifying the libraries of instructional objectives after we convert to a commitment to behavioral objectives? Will there be programmed texts that the students can use once they're through this set of programmed texts, or will they be programmed orphans? Will the competency-based movement follow through? The evaluator has to investigate such questions.

4. The Specific Problem of Teacher Evaluation

I will simply mention the fact that, apart from the problem of evaluating competency-based teacher education, there is also the problem of evaluating the problem of evaluating the products of CBTE programs, namely, competency-trained teachers. Here we run into a number of other quite different problems for the evaluator. How do you get from a conclusion about the merit for the impacted population

of having competency-based teachers to the merit of an individual teacher? There is no easy path. Even such simple questions as how one gets from competency-mastery judgments (which one makes in the course of supervising the trainees) to competency-exercise conclusions (which is what one needs in order to be sure that his competencies will be used in the classroom)? James Popham has clearly stressed the important distinction between competency-based approaches to program evaluation and competency-based approaches to individual teacher evaluation; given the present instability of data on competency of teachers as individuals, we by no means have to throw up our hands about the possibility of evaluating training programs using competency-based tests. Again, there is the question of whether we are going to evaluate teachers for minimum competency, or for purposes of developing a competency profile on which one could rely for, e.g., placing a specialist. The former is in much better shape, as the evaluation art stands, than the latter.

In short, while the competency-based approach offers us considerable opportunities for improving the present state of teacher evaluation, this mainly shows that the present state of teacher evaluation is in very bad shape, not that the competency-based approach is without its own problems.

4. Cost-free Evaluation

I promised to add a little touch of good news to the end of this list of worries, and I wish there were time to develop the good news at comparable length! But I can at least explain the general thesis that I want to propose, even if I can't get into details about the exact reasons for believing it to be true.

The general thesis is that the costing of evaluation itself is usually incorrectly done. It is usually seen as a cost on the books of either an agency or a project

or perhaps a school district, and the entry there reflects the fact that somebody has to make out a check to pay the evaluator. What doesn't show up there is the hidden side of the ledger, namely, the returns from the evaluation. Exactly the same point can be made about a fire department. If you ask the mayor of a city what the cost of the fire department is, then he or she is likely to reply by mentioning the figure that represents the amount of money paid over by the City Treasurer into the fire department account each year. Of course, the fire department isn't just a cost. The department's existence is what keeps insurance rates down in the city. The department's existence in what keeps the death rate and property damage rate down in the city. The returns from those activities of the department have to be taken into account in a correct set of social books. The same is true of evaluation. Evaluation is not a tack-on luxury, stuffed down our throats by external agencies. It is what it takes to determine the quality of one's own work. Without effective evaluation, there can be no knowledge that what has been done has been worthwhile. If, therefore, one's commitment is to worthwhile activity, then one is necessarily committed to substantial evaluation. Now, if we start drawing up a set of social books, then we have to recognize that evaluation is frequently responsible for terminating projects that are not worthwhile; and it is sometimes responsible for increasing the amounts of resources that are put into other projects of exceptional promise, thereby getting them beyond the threshold of very large scale payoffs. Evaluation, in short, should not represent a net cost to a community, such as the teacher educators. I won't say that bad evaluation don't cost more than they should; by my standards, that's an excellent ground for saying that they're bad, since they fail to outperform a critical competitor. I only say that a good evaluation is not only a good investment but the only means whereby you can recognize whether your investment is good or not. So, if I have presented a rather depressing list of activities that are involved in satisfactory evaluation, and that are not taken care of by the move

to competency-based approaches, I have at least provided some kind of a sugar coating for the pill that I am suggesting one has to swallow!

6. Conclusion

Competency-based evaluation is a means to the ends of education. Whether it becomes one more semantic tombstone in the cemetary of educational fads depends entirely on two issues; first, whether we can keep our eye on quality and quality control, and cost the standards of quality, not mere popularity; second, whether we can act like rational users or consumers and adopt on the basis of merit.